

# Total Cost of Preservation (TCP): Cost Modeling for Sustainable Services

Rev. 0.16 – 2012-04-09

Stephen Abrams

Patricia Cruse

John Kunze

*UC Curation Center, California Digital Library  
University of California, Office of the President*

Michael Mundrane

*Office of the Deputy CIO  
University of California, Berkeley*

## 1 Introduction

Information technology and resources are thoroughly integrated with, and indispensable to, today's Internet and web-based culture, commerce, science, education, and entertainment. The digital assets underpinning these activities, however, are inherently fragile with respect to ever increasing disruptive technological change. Without effective, affordable, and proactive curation management, today's digital assets will not remain viable and useful in the future. In order to achieve ubiquity of long-term preservation efforts and results, the full economic costs of preservation activities over time, the "*total cost of preservation*" (TCP), must be well understood.

TCP analysis can be applied usefully in the development of two specific price models for preservation service pricing:

- Pay-as-you-go
- Paid-up

The pay-as-you-go model is appropriate for situations where a reliable and predictable annual income stream is available to the client purchasing preservation services. When this is not the case, for example, with organizations facing irregular or boom-or-bust budgetary cycles or for grant-funded, fixed term research projects, the paid-up model may be more attractive; indeed, in many circumstances it may be the only realistic option.

## 2 Total cost of preservation

Long-term digital preservation is a complex activity, involving sophisticated technical infrastructure as well as significant human competencies, analysis, and decision making. Given the difficulties in accounting for the myriad aspects of preservation activity, and of projecting the

cost of those activities into the indefinite future, any analysis of TCP must rely on a number of fundamental assumptions. (See Appendix A for a summary of prior work on preservation cost analysis.)

## 2.1 Assumptions

The TCP analysis assumes the following:

- Only the costs pertaining to *preservation service providers* are considered. In terms of the OAIS reference model [OAIS], these are the costs associated with an *Archive*, “an organization [of people and systems]...that has accepted the responsibility to preserve information.” In particular, the preservation costs associated with the local activities of *OAIS Producers* (e.g., content creation or acquisition, reformatting, packaging, submission, etc.) are considered out of scope. On the other hand, the cost of supporting a *Producer* in making use of *Archive* functions *is* in scope.
- Individual cost components of the model can be categorized unambiguously as either *fixed*, which are incurred regardless of level of use, or *marginal*, which scale proportionally with use.
- Individual cost components can be categorized unambiguously as either *one-time* or *recurring*. One-time costs can be *annualized* over the effective lifespan of the activity or component.
- The values defined for individual cost components represent *nominal* costs, that is, the costs defined for generic instances of activities or system capacities. This is reasonable under a further assumption of a policy of uniformity of preservation *effort* rather than *outcome*. For example, on the basis of form, structure, accompanying metadata, etc., some digital content may be more inherently amenable to preservation care and will naturally receive a higher level of preservation service and outcome.
- Actions performed on preservation content, e.g., characterization, fixity, transformation, etc., are substantially automated. Thus, the main cost associated with an action is in the acquisition – or analysis, design, and development – and deployment of the implementing software, which is *independent* of the number of objects against which the action is performed.
- The size and scope of the content to be supported is known, and the cost of providing for its preservation is established, at the *beginning* of the time period under consideration. However, under the pay-as-you-go price model these costs are billed for at the *end* of the time period while under the paid-up model they are billed for at the *beginning* of the period.
- The preservation service provider can carry forward budgetary *surpluses* across fiscal year boundaries. Furthermore, these surplus funds can be *invested* at market rates, with the

earned interest further contributing to the surplus.

- The models are ultimately *revenue neutral*, that is, there should be no surplus (or deficit) funds remaining at the end of the period under consideration.
- Values can be determined for various *annual adjustment factors*, such as inflation, investment rate of return, cost of living adjustments (COLA), merit pay raises, and changes to the per-unit cost of preservation, which, although held *constant* over the full period of TCP consideration, are nevertheless reliably *predictive* of long-term economic, technological, and organizational trends.

The last assumption, although the mainstay of the standard economic forecasting technique of *discounted cash flow* (DCF) analysis [Damodaran], can be problematic over extended time periods. Concerns over the efficacy of DCF-based modeling are addressed in Section § 3.

## 2.2 TCP analysis

The TCP analysis encompasses the full economic costs associated with the long-term preservation of digital assets, although the resulting price models can be easily customized to deal only with various subsets of those costs as a matter of local policy. The individual cost components incorporated in the analysis are reflective of the following 11 high-level categories abstracted from the environment in which preservation activities occur (see Table 1):

Preservation activities are embodied in an archival

1. **System**; composed of various
2. **Services** supporting necessary and desirable functions; running on
3. **Servers**; designed, deployed, maintained, enhanced, and utilized by
4. **Staff**; in support of content
5. **Producers**; who use
6. **Workflows** to submit instances of
7. **Content Types**; which occupy
8. **Storage**; and are subject to ongoing
9. **Monitoring**; and periodic
10. **Interventions**; all subject to appropriate managerial
11. **Oversight**.

Table 1 – Preservation environment

Each of the preceding *italicized* items is a cost component of the TCP analysis. To simplify that

analysis, the *System* component encompasses the baseline technical infrastructure and environment in which preservation activities are embodied, and in particular, subsumes the *Services* that support necessary technical functions and the *Servers* on which those *Services* are deployed. Similarly, *Staff* costs are subsumed by the other individual high-level cost component terms.

The *Content Type* component encompasses all analysis, planning, and software acquisition or development specific to supporting given class of digital content.

The *Monitoring* component encompasses technology watch for incipient obsolescence and other risk factors that may impinge on long-term usability; routine pro-active monitoring of the *System* (and its underlying *Services* and *Servers*) and *Storage* for availability and responsiveness, on the other hand, are subsumed under the *System* and *Storage* components respectively.

The *Intervention* component encompasses all non-periodic or unpredictable preservation activities such as format migrations, responses to rights challenges, disaster recovery efforts, etc., storage media refresh and migration are subsumed under the *Storage* component.

The TCP analysis represents the total cost of preservation to the preservation service provider as follows:

$$TCP = A + n \cdot P + m \cdot W + \ell \cdot C + k \cdot S + j \cdot M + i \cdot V + O \quad (1)$$

<i>TCP</i>	Total cost of preservation for all <i>Producers</i> .
<i>A</i>	Fixed cost of the baseline archival <i>System</i> .
<i>n</i>	Number of content <i>Producers</i> .
<i>P</i>	Unit cost of supporting a <i>Producer</i> .
<i>m</i>	Number of submission <i>Workflows</i> .
<i>W</i>	Unit cost of supporting a <i>Workflow</i> .
<i>ℓ</i>	Number of <i>Content Types</i> .
<i>C</i>	Unit cost of supporting a <i>Content Type</i> .
<i>k</i>	Number of units of preservation <i>Storage</i> .
<i>S</i>	Unit cost of <i>Storage</i> .
<i>j</i>	Number of preservation <i>Monitoring</i> activities.
<i>M</i>	Unit cost of a <i>Monitoring</i> activity.
<i>i</i>	Number of preservation <i>Interventions</i> .
<i>V</i>	Unit cost of an <i>Intervention</i> .
<i>O</i>	Fixed cost of administrative and managerial <i>Oversight</i> .

Table 2 – TCP notation

The baseline archival *System* and administrative/managerial *Oversight* are considered fixed costs, therefore they are represented in Equation (1) by single terms,  $A$  and  $O$ , respectively. All other components are marginal costs, each representing by two terms: a unit cost and the number of units allocated or consumed.

The TCP analysis can be used as the basis for partial cost-recovery by removing or zeroing-out particular cost components that are subsidized as a matter of local policy. The specific pay-as-you-go and paid-up price models are based on a derivation of the preservation costs attributable to a given content *Producer*.

### 2.3 Pay-as-you-go price model

The pay-as-you-go price model is based on an annual billing cycle,  $t = 0, 1, 2, \dots$ . Certain cost components – *System*, *Workflow*, *Content Type*, *Monitoring*, *Intervention*, and *Oversight* – are considered “common goods”; that is, they are applicable and beneficial to all *Producers* equally, and as such, are properly apportioned across all *Producers*. The number of units of *Storage*, on the other hand, is specific to a single *Producer*. The annual pay-as-you-go price for a given *Producer* is:

$$G = \frac{A + m \cdot W + \ell \cdot C + j \cdot M + i \cdot V + O}{n} + P + k_p \cdot S \quad (2)$$

$G$  Pay-as-you-go price.

$k_p$  Number of units of preservation *Storage* attributable to a given *Producer*.

Table 3 – Pay-as-you-go price model notation

The operation of the pay-as-you-go model is illustrated in the cash flow diagram [Neftci] in Figure 1. (Values above the horizontal time axis represent *income* or price; those below the axis represent *expenses* or costs.) A series of equal costs  $G$  are incurred at the *end* of each service year. To achieve revenue neutrality these are matched with corresponding prices.

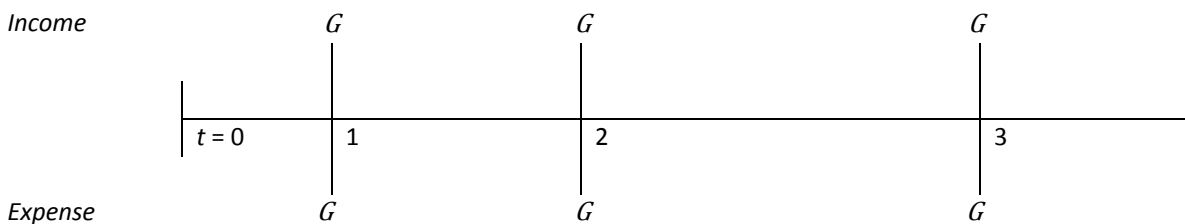


Figure 1 – Pay-as-you-go price model cash flow diagram

The long-term cost to a *Producer* under the pay-as-you-go model is found by summing together the annual payments  $G$  over a period of  $T$  years.

$$G(T) = G + G + \cdots + G \quad (3)$$

$$G(T) = G \cdot T \quad (4)$$

$G(T)$  Cumulative pay-as-you-go price over  $T$  years.

$G$  Annual pay-as-you-go price.

$T$  Fixed term of  $T$  years.

Table 4 – Pay-as-you-go price model notation

Unfortunately, this cumulative price increases linearly as a function of time and, most alarmingly, as  $T$  approaches “forever” ( $T \rightarrow \infty$ ), the total price also approaches infinity.

$$G(\infty) = \infty \quad (5)$$

However, recalling the assumption of an annual decrease in the aggregate cost of providing preservation service,  $d$ , the series of annual costs and prices decreases uniformly, as illustrated in Figure 2.

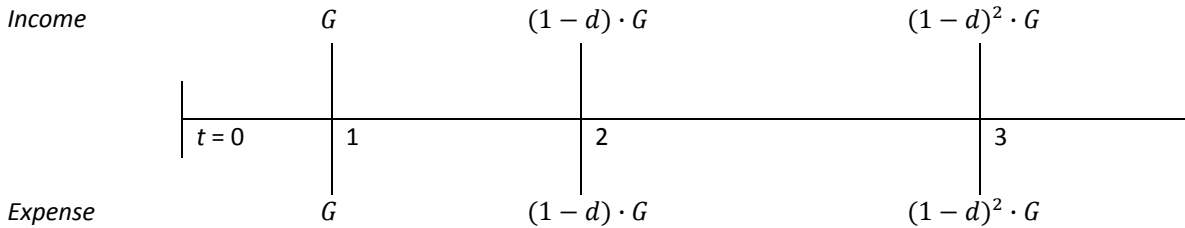


Figure 2 – Discounted pay-as-you-go price model cash flow diagram

The size of the annual decrease is proportional to  $1-d$ , which compounds over time. Again, the long-term cost of is found by summing together the annual payments over  $T$  years.

$$G(T, d) = G + G \cdot (1 - d) + G \cdot (1 - d)^2 + \cdots + G \cdot (1 - d)^{T-1} \quad (6)$$

$$G(T, d) = G \cdot \frac{1 - (1 - d)^T}{d} \quad (7)$$

$$G(\infty, d) = \frac{G}{d} \quad (8)$$

$G(T, d)$  Cumulative discounted pay-as-you-go price over  $T$  years.

$d$  Annual percentage rate of decrease in aggregate preservation service cost (as a decimal, for example, 0.1 for a 10% annual decrease).

Table 5 – Pay-as-you-go price model notation (continued)

Mathematically, Equations (7) and (8) are the solutions of finite and infinite *power series*

respectively, which are solvable only when the rate of decrease in aggregate preservation service cost,  $1-d$ , is less than 1. For details of the derivation of these equations, see Appendices B and C. For the derivation of the value of  $d$  see Appendix D.

### 2.3 Paid-up price model

The paid-up price model is based on a one-time price, paid at the beginning of the period under consideration, that is sufficient to finance all subsequent preservation activities over that period. This period can be either a fixed term  $T$  or “forever,” that is, for  $T = \infty$ . In the relevant cash flow diagram in Figure 3, the one-time payment  $F$  is paid by the *Producer* at the beginning of year 1 ( $t = 0$ ). The subsequent income terms are the result of the investment return  $r$  applied to each year’s surplus funds. Note that  $r$  is the *nominal* annual percentage rate (APR) of investment return. The *effective* annual rate,  $e$ , incorporating the effect of monthly compounding, is [Tucker]:

$$e = \left(1 + \frac{r}{12}\right)^{12} - 1 \quad (9)$$

The pay-as-you-go cost  $G$  is scaled by a factor of  $e/r$  to account for the fact that the service provider incurs costs *throughout* the service year but these are not billed and paid for until the *end* of the year.

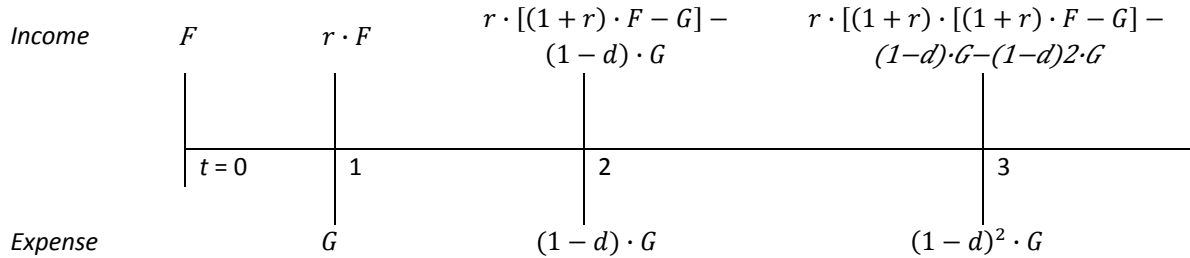


Figure 3 – Paid-up price model cash flow diagram

The paid-up price  $F$  is derived in Appendix E.

$$F(T, d, r) = G \cdot \frac{e}{r} \cdot \frac{1}{1+e} + G \cdot \frac{e}{r} \cdot \frac{(1-d)}{(1+e)^2} + G \cdot \frac{e}{r} \cdot \frac{(1-d)^2}{(1+e)^3} + \dots + G \cdot \frac{e}{r} \cdot \frac{(1-d)^{T-1}}{(1+e)^T} \quad (10)$$

$$F(T, d, r) = G \cdot \frac{e}{r} \cdot \frac{(1+e)^T - (1-d)^T}{(1+e)^T \cdot (e+d)} \quad (11)$$

$$F(\infty, d, r) = G \cdot \frac{e}{r} \cdot \frac{1}{e+d} \quad (12)$$

$F(T,d,r)$	Paid-up price for $T$ years.
$r$	The nominal annual percentage rate (APR) of investment return (as a decimal, for example, 0.02 for 2%.
$e$	The effective annual percentage rate of investment return (as a decimal).

Table 6 – Paid-up price model notation

Equations (11) and (12) are also power series and are solvable only when the discount factor is less than the effective investment return,  $1-d < 1+e$ . For details of the derivation of these equations, see Appendix E.

It is desirable for a *Producer* to be able to switch easily between a pay-as-you-go and paid-up basis. The relevant factor in this transition is the “coefficient of permanence,” that is, the one-time premium, or multiplier, of the annual pay-as-you-go price,  $G$ , that one must incur in order to achieve paid-up permanence.

$$\varphi = \frac{e}{r \cdot (e + d)} \quad (13)$$

$\varphi$	Coefficient of permanence; the one-time multiplier of the pay-as-you-go price necessary to achieve paid-up permanence.
-----------	--

Table 7 – Paid-up model notation (continued)

Thus, a *Producer* currently paying  $G$  on an annual basis can “trade up” to service permanence with a one-time payment of  $F = \frac{G \cdot e}{r \cdot (e + d)}$ .

The paid-up model is designed to be ultimately revenue neutral, that is, with no surplus (or deficit) funds remaining at the end of the term. This can be illustrated in the following example for a ten year term ( $T = 10$ ), a pay-as-you-go price  $G$  of \$650 (for 1 TB of storage), a nominal annual investment rate of return of 2% ( $r = 0.02$ ), and an annual decrease in the aggregate cost of preservation service of 10% ( $d = 0.1$ ). From Equation (11), this leads to a paid-up price  $F$  of \$4,764.14. Income in years 1,2,3,... is the result of the investment return on the surplus remaining at the end of the prior year.

Year ( $T$ )	Income	Expense	Surplus
0	\$ 4,764.14	\$ 0.00	\$ 4,764.14
1	\$ 96.16	\$ 655.99	\$ 4,204.31
2	\$ 84.86	\$ 623.19	\$ 3,665.98
3	\$ 73.00	\$ 592.03	\$ 3,147.94



4	\$ 63.54	\$ 562.43	\$ 2,649.05
5	\$ 53.47	\$ 534.31	\$ 2,168.21
6	\$ 43.76	\$ 507.59	\$ 1,704.38
7	\$ 34.40	\$ 482.21	\$ 1,256.57
8	\$ 25.36	\$ 458.10	\$ 823.83
9	\$ 16.63	\$ 435.20	\$ 405.26
10	\$ 8.18	\$ 413.44	\$ 0.00

Table 8 – 10 year term paid-up price model example

The relationship between the pay-as-you-go and paid-up models is illustrated in Figure 4. The cumulative and discounted pay-as-you-go prices and the paid-up price,  $G(T)$ ,  $G(T,d)$ ,  $F(T,d,r)$ , are shown as the red, orange, and green curves, respectively. The cumulative pay-as-you-go and paid-up prices for “forever,”  $G(\infty,d)$  and  $F(\infty,d,r)$ , are shown in magenta and blue, respectively.

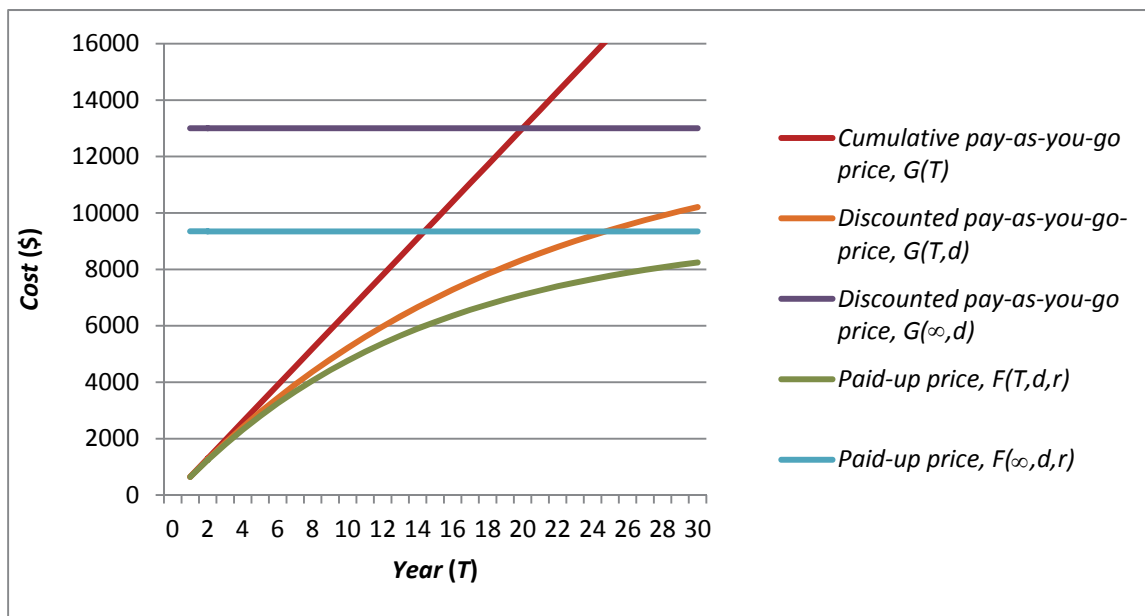


Figure 4 – Relationship between pay-as-you-go and paid-up price models

The difference between the cumulative (red) and discounted (orange) pay-as-you-go prices is accounted for by the discounting factor  $d$ , the assumed annual decrease in the cost of providing preservation service. The difference between the fixed term pay-as-you-go (orange) and paid-up (green) prices, as well as the “forever” prices (magenta and blue), is accounted for by  $r$ , the assumed annual investment rate of return.

The curve for the fixed term paid-up price characteristically approaches the forever price asymptotically so long as the investment rate of return and annual decrease in aggregate

preservation service cost are greater than 0 ( $r > 0$  and  $d > 0$ ), or if the aggregate cost *increases* (i.e.,  $d < 0$ ), when the investment gain is greater than the service cost increase ( $r > |d|$ ).

## 2.4 Rate of change of preservation service cost

The variable  $d$  in Equations (7), (8), (11), and (12) is a composite value or function that sums the individual component-specific changes in quantity and unit cost, weighted by the component's proportionate contribution to the overall cost (see Appendix E).

## 3 Predicative reliability

Many of the assumptions built into the TCP model may be problematic over significant time spans, particularly the assumption that the various annual adjustment factors – investment rate, rate of decrease of unit cost, inflation, COLA, etc. – are constant over time [Rosenthal].

Several avenues are available to ameliorate these difficulties.

### 3.1 Model the risk

The values for  $r$  and  $d$  can be rounded up (or down) as appropriate to represent a fixed risk “premium” for the future uncertainty of financial and technological trends. A better approach is to add an additional risk component  $R$  to the pay-as-you-go Equation (2).

$$G = \frac{A + m \cdot W + \ell \cdot C + j \cdot M + i \cdot V + O + R}{n} + P + k_P \cdot S \quad (14)$$

$R$ Risk cost posed by uncertainty over the future values of $r$ and $d$ .
--

Table 9 – Risk notation

The growing influence of  $R$  over time, reflective of increasing uncertainty, can be controlled by setting the appropriate value for  $\omega_R$  and  $d_R$ , the weighting factor and expected annual rate of change in risk used in Equation (14) for  $d$ , the rate of change for the aggregate cost of preservation in Equations (7), (8), (11), and (12).

Insurers protect themselves from truly unexpectedly large payouts by purchasing reinsurance: in essence, insurance against excessive claims [Reinsurance]. The additional cost introduced into the pricing equations through the risk factor  $R$  can be seen as representing a reinsurance “premium” that accumulates over time in anticipation of meeting some future unexpected need. This is also similar to the idea of a “rainy day fund” that governments fund during periods of budgetary surplus in anticipation of future shortfalls [Rainy].

### 3.2 Recalibration

Rather than keeping key model variables,  $F$ ,  $G$ ,  $r$ , and  $d$ , constant over time, they can be recalibrated periodically on the basis of cumulative experience and current knowledge. Recalibrated prices would apply for prospective *Producers* only; retrospective *Producers* would remain “locked in” at the price in force at the time they established their relationship with the service provider.

### 3.3 Hybrid pricing

The hybrid approach to pricing distinguishes between those cost components that are considered easy to quantify and forecast, and those that aren’t. Paid-up pricing is applied only to the former; activities associated with the latter are performed, if it all, on a pay-as-you-go basis.

For example preservation *Interventions*, future actions taken in response to future, possibly unanticipated, conditions, may be considered to fall into the difficult-to-quantify category. The paid-up price would therefore only incorporate the costs for preservation activities that can be performed at the time of content ingest via *Workflows* as the result of *Content Type* analysis, such as the creation of normalized or desiccated [WAS, Desiccation] derivative forms of the contributed content.

In a more extreme case, the baseline paid-up preservation function could be restricted to bit-level preservation only. This service mode requires only three cost components: the *System*, *Storage*, and *Producer*; support for *Workflow* is not necessary as the single required workflow is assumed to be subsumed under the *System*, while *Content Types*, *Monitoring*, and *Interventions* are not necessary since only bit-level activities are encompassed at this service level.

### 3.4 Shorten the term

Uncertainty in the predictive reliability of the values chosen for  $r$  and  $d$  is proportion to the length of time under consideration; the longer the term, the greater the uncertainty. With this in mind, it may be prudent to shorten the term to one more amenable to reliable forecasting, with the possibility for subsequent “renewal” of preservation service. For some bodies of content a fixed term – 10 year? 20 years? – may in fact be an appropriate initial term. Short to mid-term preservation provides an opportunity for the content to prove its value, as evidenced by the commitment of a curatorial advocate to pay for its subsequent preservation.

### 3.5 Stochastic modeling

Rosenthal suggests using more sophisticated economic techniques based on Monte Carlo simulation to model potential future scenarios stochastically [Rosenthal].

## Appendix A Prior work

A number of international efforts have studied the question of long-term preservation costs; most notably, the Nationaal Archief of the Netherlands in 2005 [Archief]; the LIFE<sup>2</sup> project (Life Cycle Information for E-Literature) work on a Generic Preservation Model (GPM) in 2008 [Ayrís, Hole, LIFE]; the KRDS project (Keeping Research Data Safe) in 2010 [Beagrie, KRDS]; the Princeton DataSpace initiative in 2010 [Goldstein]; and the Danish National Archives and Royal Library CMDP (Cost Model for Digital Preservation) in 2011 [Kejser, CMDP]. (See [Zeller] for a summary of these and other relevant activities.)

All of these models analyze preservation costs throughout the full lifecycle of preserved assets. Both the Nationaal Archief and LIFE project work employ very fine-grained analysis of cost components and are based on representative *actual* costs rather than the *nominal* costs employed by TCP. The LIFE model includes cost components for content creation and acquisition, which are considered out of scope in the TCP analysis. Furthermore, LIFE scales preservation action costs by the number of expected objects. TCP assumes that there is no per-object marginal cost; instead, marginal costs are associated only with the various *Content Types* of which objects are members. The KRDS model is specific to the research data lifecycle, but its findings appear to be applicable to other contexts. Unlike the the Nationaal Archief and LIFE models, KRDS assumes a discounting function that annually decreases the aggregate cost of preservation service. The DataSpace model is also based on a discounting function, but its analysis covers only the costs associated with preservation storage, which are defined on a pay once, store forever (POSF) basis. The OAIS activity-based CMDP work concentrates on post facto measuring of preservation costs, rather than on forecasting, although it is possible using the framework. While DataSpace is explicitly concerned with “forever” pricing, neither it nor any of the other models assume the benefit of an annual investment return in offsetting a portion of ongoing costs. Most, if not all, of the individual cost components of these models can be mapped to OAIS environmental and functional entities [OAIS] or TRAC criteria [TRAC], facilitating common points of reference and comparison.

A number of institutions, both commercial and non-profit/academic, now offer long-term preservation services. (Or at least persistent storage; the important distinction between the two is not always apparent from the description of these service offerings.) Carbonite is representative of commercial preservation service offerings [Carbonite]. Its cost is \$599/year for up to 500 GB, with further increments of 100 GB available for \$89/year. The LifeTime Library at the University of North Carolina offers students permanent storage “and associated services” for 250 GB, apparently with no associated fees [UNC]. The USC Digital Repository offers a paid up license for 20 years of preservation service for \$1000/TB [USC]. In general, however, little information is available explaining the basis of their business models.

## Appendix B Power series

The solution for a general power series,  $p(x) = \sum_{i=0}^{T-1} a \cdot x^i$ , can be derived as follows:

$$p(x) = a + a \cdot x + a \cdot x^2 + \dots + a \cdot x^{T-1} \quad (15)$$

Equation (15) can be scaled by a factor of  $x$ :

$$x \cdot p(x) = a \cdot x + a \cdot x^2 + \dots + a \cdot x^{T-1} + a \cdot x^T \quad (16)$$

Subtracting Equation (16) from (15) yields:

$$p(x) - x \cdot p(x) = a + (a \cdot x - a \cdot x) + \dots + (a \cdot x^{T-1} - a \cdot x^{T-1}) - a \cdot x^T \quad (17)$$

$$p(x) - x \cdot p(x) = a - a \cdot x^T \quad (18)$$

$$p(x) \cdot (1 - x) = a \cdot (1 - x^T) \quad (19)$$

Thus, dividing Equation (19) by  $(1-x)$ :

$$p(x) = a \cdot \frac{1 - x^T}{1 - x} \quad \text{where } x \neq 1 \quad \blacksquare \quad (20)$$

The equation for  $p(x)$  as an *infinite* series is derived by observing that  $x^T$  in Equation (20) approaches 0 as  $T$  approaches  $\infty$ , so long as  $0 \leq x < 1$ .

$$\lim_{T \rightarrow \infty} p(x) = \frac{a - 0}{1 - x} \quad (21)$$

$$\lim_{T \rightarrow \infty} p(x) = \frac{a}{1 - x} \quad \text{where } 0 \leq x < 1 \quad \blacksquare \quad (22)$$

Various other equations in the TCP models are derived by substituting the appropriate terms for  $a$  and  $x$  in Equations (20) and (22).

## Appendix C Pay-as-you-go price model

The cumulative discounted pay-as-you-go price over  $T$  years is based on the sum of a geometrically decreasing series of payments, each proportionally discounted from  $G$  by a factor of  $1-d$  (see Figure 2):

$$G(T, d) = G + G \cdot (1-d) + G \cdot (1-d)^2 + \dots + G \cdot (1-d)^{T-1} \quad (23)$$

The solution for Equation (23) is simply the general solution for a finite power series given by Equation (20) with  $G$  substituted for  $a$  and  $1-d$  for  $x$ .

$$G(T, d) = a \cdot \frac{1 - x^T}{1 - x} \quad \text{where } a = G \text{ and } x = 1-d \quad (24)$$

Substituting for  $a$  and  $x$  and simplifying produces the cumulative discounted pay-as-you-go price over  $T$  years.

$$G(T, d) = G \cdot \frac{1 - (1-d)^T}{d} \quad \text{where } d \neq 0 \quad \blacksquare \quad (25)$$

Likewise, the finite power series also has a solution for the case where time is arbitrarily large as long as it satisfies the convergence criterion,  $0 \leq x < 1$ .

$$G(\infty, d) = \frac{a}{1 - x} \quad \text{where } a = G \text{ and } x = 1-d \quad (26)$$

Substituting in for  $a$  and  $x$  and simplifying produces the cumulative discounted pay-as-you-go price for “forever.”

$$G(\infty, d) = \frac{G}{d} \quad \text{where } 0 \leq 1-d < 1 \quad \blacksquare \quad (27)$$

## Appendix D Discount factor

The variable  $d$  in Equations (7), (8), (11), and (12) is a composite value or function that sums the individual component-specific changes in quantity and unit cost, weighted by the component's proportionate contribution to the overall cost (see Appendix E).

$$d = \omega_A \cdot d_A + \omega_W \cdot (d_m \cdot d_W) + \omega_C \cdot (d_\ell \cdot d_C) + \omega_M \cdot (d_j \cdot d_M) + \omega_V \cdot (d_i \cdot d_V) + \omega_S \cdot (d_k \cdot d_S) + \omega_O \cdot d_O \quad (28)$$

where

$$\omega_A = \frac{A}{n \cdot G} \quad (29)$$

$$\omega_W = \frac{m \cdot W}{n \cdot G} \quad (30)$$

$$\omega_C = \frac{\ell \cdot C}{n \cdot G} \quad (31)$$

$$\omega_M = \frac{j \cdot M}{n \cdot G} \quad (32)$$

$$\omega_V = \frac{i \cdot V}{n \cdot G} \quad (33)$$

$$\omega_S = \frac{k \cdot S}{G} \quad (34)$$

$$\omega_O = \frac{A}{n \cdot G} \quad (35)$$

$d$	Annual percentage rate of change in aggregate preservation service cost.
$\omega_A$	Weighting factor for the <i>System</i> component.
$d_A$	Annual percentage rate of change in fixed cost of the archival <i>System</i> .
$\omega_W$	Weighting factor for the <i>Workflow</i> component.
$d_m$	Annual percentage rate of change in the number of supported <i>Workflows</i> .
$d_W$	Annual percentage rate of change in the nominal unit cost of a <i>Workflow</i> .
$\omega_C$	Weighting factor for the <i>Content Type</i> component.
$d_\ell$	Annual percentage rate of change in the number of supported <i>Content Types</i> .
$d_C$	Annual percentage rate of change in the nominal unit cost of a <i>Content Type</i> .
$\omega_M$	Weighting factor for the <i>Monitoring</i> component.
$d_j$	Annual percentage rate of change in the number of supported <i>Monitoring</i> activities.
$d_M$	Annual percentage rate of change in the nominal unit cost of a <i>Monitoring</i> activity.
$\omega_V$	Weighting factor for the <i>Intervention</i> component.

$d_i$	Annual percentage rate of change in the number of supported <i>Interventions</i> .
$d_v$	Annual percentage rate of change in the nominal unit cost of an <i>Intervention</i> .
$\omega_s$	Weighting factor for the <i>Storage</i> component.
$d_k$	Annual percentage rate of change in the number of <i>Storage</i> units consumed.
$d_s$	Annual percentage rate of change in the nominal cost of a unit of <i>Storage</i> .
$\omega_o$	Weighting factor for the <i>Oversight</i> component.
$d_o$	Annual percentage rate of change in the number cost of a unit of <i>Oversight</i> .

Table 11 – Change in preservation service cost notation

Since the *System* and *Oversight* are fixed costs, the numerators of Equations (29) and (35) only have the unit cost term; all other weighting factors have both number of units and unit cost terms. Since the costs for *System*, *Workflows*, *Content Types*, *Monitoring*, *Interventions*, and *Oversight* are common goods, they are apportioned across all  $n$  *Producers*, as reflected in the denominators of Equations (29) – (33) and (35); *Storage* is always specific to a given *Producer*, so there is no  $n$  term in Equation (34).

Note that *all* of the discount factors are *real* rates of change, incorporating the effects, if any, of inflation.



## Appendix E Paid-up price model

Like the pay-as-you-go price, the paid-up price over  $T$  years is based on the sum of a geometrically decreasing series of payments (see Figure 3). The present value  $F$  of the future series of payments using the base cost  $G$  declining annually by  $1-d$ , and a *nominal* investment return  $r$  compounded monthly is:

$$F(T, d, r) = G \cdot \frac{e}{r} \cdot \frac{1}{1+e} + G \cdot \frac{e}{r} \cdot \frac{1-d}{(1+e)^2} + G \cdot \frac{e}{r} \cdot \frac{(1-d)^2}{(1+e)^3} + \dots + G \cdot \frac{e}{r} \cdot \frac{(1-d)^{T-1}}{(1+e)^T} \quad (36)$$

The *effective* investment return  $e$ , incorporating the effect of monthly compounding, is [Tucker]:

$$1 + e = \left(1 + \frac{r}{12}\right)^{12} \quad (37)$$

Each declining term,  $G \cdot (1-d)^t$ , representing a future payment is converted to its present value by adjusting for the monthly-compounded time value of money, that is, scaling by  $\frac{e}{r} \cdot (1+e)^{t+1}$ , and the results are summed to produce the total present value. Rearranging the terms of Equation (36) gives:

$$F(T, d, r) = \frac{G}{1+e} \cdot \frac{e}{r} \cdot \left(1 + \frac{1-d}{1+e} + \frac{(1-d)^2}{(1+e)^2} + \dots + \frac{(1-d)^{T-1}}{(1+e)^{T-1}}\right) \quad (38)$$

The solution to Equation (38), is simply the general solution to the power series given by Equation (20) where  $a$  is  $\frac{G}{1+e} \cdot \frac{e}{r}$  and  $x$  is  $\frac{1-d}{1+e}$ .

$$F(T, d, r) = a \cdot \frac{1-x^T}{1-x} \quad \text{where } a = \frac{G}{1+e} \cdot \frac{e}{r} \text{ and } x = \frac{1-d}{1+e} \quad (39)$$

Substituting for  $a$  and  $x$  and simplifying produces the paid-up price for  $T$  years.

$$F(T, d, r) = G \cdot \frac{e}{r} \cdot \frac{(1+e)^T - (1-d)^T}{(1+e)^T \cdot (e+d)} \quad \blacksquare \quad (40)$$

Likewise, the finite power series also has a solution for the case where time is large as long as it satisfies the convergence criterion,  $0 \leq x < 1$ .

$$F(\infty, d, r) = \frac{a}{1-x} \quad \text{where } a = \frac{G}{1+e} \cdot \frac{e}{r} \text{ and } x = \frac{1-d}{1+e} \quad (41)$$

Substituting for  $a$  and  $x$  and simplifying produces the paid-up price for “forever.”

$$F(\infty, d, r) = \frac{e}{r} \cdot \frac{G}{e + d} \quad \text{where } 0 \leq \frac{1 - d}{1 + e} < 1 \quad \blacksquare \quad (42)$$

## References

- [Archief] Nationaal Archief (2005), *Costs of Digital Preservation*, <<http://www.nationaalarchief.nl/sites/default/files/docs/kennisbank/codpv1.pdf>>, accessed 2011-09-02. See also, <[http://www.nationaalarchief.nl/sites/default/files/docs/kennisbank/kostenmodel\\_in\\_excel\\_versie\\_1.2\\_final.xls](http://www.nationaalarchief.nl/sites/default/files/docs/kennisbank/kostenmodel_in_excel_versie_1.2_final.xls)>, accessed 2011-09-02.
- [Ayriss] Ayriss, Paul, Richard Davies, Rory McLeod, Rui Miao, Helen Shenton, and Paul Wheatley (2008), *LIFE2 Final Project Report*, <<http://eprints.ucl.ac.uk/11758>>.
- [Beagrie] Beagrie, Neil, Brian Lavoie, and Matthew Woollard (2010), *Keeping Research Data Safe 2* <<http://www.jisc.ac.uk/media/documents/publications/reports/2010/keepingresearchdatasafe2.pdf>>.
- [Carbonite] Carbonite, Inc. (2011), *Online Backup – Data & File Backup Software | Carbonite* <<http://www.carbonite.com/en/default.aspx>>.
- [CMDP] Danish National Archives/Danish Royal Library (2011), *Cost Model for Digital Preservation (CMDP)* <<http://www.costmodelfordigitalpreservation.dk/>>, accessed 2011-09-02. See also, <[http://www.costmodelfordigitalpreservation.dk/contact/cmdp-1---preservation-planning-and-digital-migrations/documentation/Kostenmodel\\_in\\_Excel\\_versie\\_1.2\\_%28final%29.xls](http://www.costmodelfordigitalpreservation.dk/contact/cmdp-1---preservation-planning-and-digital-migrations/documentation/Kostenmodel_in_Excel_versie_1.2_%28final%29.xls)> and <[http://www.costmodelfordigitalpreservation.dk/contact/cmdp-2---ingest-and-archival-storage/documentation/Tool\\_Version\\_2.xls](http://www.costmodelfordigitalpreservation.dk/contact/cmdp-2---ingest-and-archival-storage/documentation/Tool_Version_2.xls)>.
- [Damodaran] Damodaran, Aswath (2002), *Investment Valuation: Tools and Techniques for Determining the Value of Any Asset* (2nd ed.; New York: Wiley).
- [Dappert] Dappert, Angela (2011), "Is it possible to predict the cost of longterm digital preservation?" Open Planets Foundation blog, <<http://www.openplanetsfoundation.org/blogs/2011-08-03-ist-possible-predict-cost-longterm-digital-preservation>>, accessed 2011-09-02.
- [Desiccation] Phillips, Mark Edward, and Daniel Gelaw Alemneh (2011), "Data Desiccation: Facilitating long-term access, use, and reuse of ETDs," *Proceedings of the 14th International Symposium on Electronic Theses and Dissertations*, Cape Town, 2011 <<http://digital.library.unt.edu/ark:/67531/metadc67625/>>.
- [Goldstein] Goldstein, Serge J., and Mark Ratliff (2010), *DataSpace: A Funding and Operational Model for Long-Term Preservation and Sharing of Research Data*, <<http://arks.princeton.edu/ark:/88435/dsp01w6634361k>>, accessed 2011-09-02.
- [Hole] Hole, Brian, Li Lin, Patrick McCann, and Paul Wheatley (2010), "LIFE<sup>3</sup>: A predictive costing tool for digital collections," *7th International Conference on Preservation of Digital Objects (iPRES2010)*, Vienna, September 19-24, 2010 <[http://www.life.ac.uk/3/docs/lpres2010\\_life3\\_submitted.pdf](http://www.life.ac.uk/3/docs/lpres2010_life3_submitted.pdf)>, accessed 2011-12-08.
- [Keijser] Keijser, Ulla Bøgvad, Anders Bo Nielsen, and Alex Thirifays (2011), "Cost model for digital preservation: Cost of digital migration," *International Journal of Digital Curation* 6(1): 255-267 <<http://www.ijdc.net/index.php/ijdc/article/viewFile/177/246>>, accessed 2011-12-08.
- [Kosten] Nationaal Archief (2011), *Kennisbank digital bewaren: Kosten* <<http://www.nationaalarchief.nl/kennisbanken/digitaal->

- [bewarenauteur=All&onderwerp=28>](#), accessed 2011-09-02.
- [KRDS] Charles Beagrie (2011), *Keeping Research Data Safe (KRDS)* <<http://www.beagrie.com/krds.php>>, accessed 2011-09-02.
- [LIFE] LIFE project (2011), *LIFE: Life Cycle Information for E-Literature* <<http://www.life.ac.uk/>>.
- [Neftci] Neftci, Salih N.. (2008), *Principles of Financial Engineering* (2nd ed.; Amsterdam: Elsevier) <<http://www.bizknowledge.info/My-books/Academic-Press-Principles-of-Financial-Engineering-2nd.pdf>>.
- [OAIS] ISO 14721 (2003), *Space data and information transfer systems – Open archival information system – Reference model*. See also, CCSDS 650.0-B-1 (2002), Blue Book <<http://public.ccsds.org/publications/archive/650x0b1.pdf>>.
- [Presto] PrestoPRIME (2011), *Storage Planning Tool*, version 0.1.4 <<http://prestoprime.it-innovation.soton.ac.uk/planning-tool/>>, accessed 2011-09-02.
- [Rainy] Brookings Institute (2010), *State and Local Tax Policy: what are rainy day funds and how do they work?* <<http://www.taxpolicycenter.org/briefing-book/state-local/fiscal/rainy-day.cfm>>, accessed 2012-01-20.
- [Reinsurance] Reinsurance Association of America (2007), *Introduction to Property and Casualty Reinsurance* <<http://www.reinsurance.org/files/public/07FundamentalsandGlossary1.pdf>>, accessed 2012-01-19.
- [Rosenthal] Rosenthal, David (2011), “Modeling the economics of long-term storage,” *dshr’s blog*, September 27, 2011 <<http://blog.dshr.org/2011/09/modeling-economics-of-long-term-storage.html>>.
- [TRAC] OCLC/Center for Research Libraries (2007), *Trusted Repository Audit & Certification: Criteria and Checklist* <[www.crl.edu/sites/default/files/attachments/pages/trac\\_0.pdf](http://www.crl.edu/sites/default/files/attachments/pages/trac_0.pdf)>.
- [Tucker] Tucker, William R. (2000), *Effective interest rate (EIR)*, Bankakademie Micro Banking Competence Center. <<http://www.docstoc.com/docs/3649810/William-R-Tucker-Effective-Interest-Rate-EIR-EIR-Periodic-Interest>>, accessed 2012-04-06.
- [UNC] University of North Carolina (2011), *Life Time Library thought to be the first nationwide*, September 15, 2011 <<http://sil.unc.edu/news/2011/SILS-lifetime-library>>.
- [USC] University of Southern California (2011), *USC Digital Repository: Pricing* <<http://repository.usc.edu/pricing/>>, accessed 2011-12-08.
- [WAS] Kunze, John (2005), “Web Archiving Service (WAS): An NDIIPP Collaboration: LC/CDL/UNT/NYU,” *DLF Fall Forum*, Charlottesville, November 7-9, 2005 <<http://old.diglib.org/forums/fall2005/presentations/kunze-2005-11.pdf>>.
- [Wheatley] Wheatley, Paul, P. Bright, and Rory McLeod (2008), *LIFE Generic Preservation Model v.1.4* <<http://discovery.ucl.ac.uk/14128/1/14128.xls>>.
- [Zeller] Zeller, Jean-Daniel (2010), “Cost of digital archiving: Is there a universal model?” *8th European Conference on Digital Archiving*, Geneva, April 28-30, 2010 <[http://regarddejanus.files.wordpress.com/2010/05/costsdigitalarchiving- jdz\\_eca2010.pdf](http://regarddejanus.files.wordpress.com/2010/05/costsdigitalarchiving- jdz_eca2010.pdf)>, accessed 2011-09-02.